ST5209/X Assignment 5

Due 21 Apr, 11.59pm

Set up

1. Make sure you have the following installed on your system: LaTeX, R4.2.2+, RStudio 2023.12+, and Quarto 1.3.450+.
2. Pull changes from the course repo.

Submission

1. Render the document to get a .pdf printout.
2. Submit both the .qmd and .pdf files to Canvas.

1. Modeling with ARIMA (Q9.7 in Hyndman & Athanasopoulos)

Consider `aus_airpassengers`, the total number of passengers (in millions) from Australian air carriers for the period 1970-2016.

a. Use `ARIMA()` to find an appropriate ARIMA model. What model was selected. Check that the residuals look like white noise. Plot forecasts for the next 10 periods.

b. Write the model in terms of the backshift operator.

c. Plot forecasts from an ARIMA(0,1,0) model with drift and compare these to part a.

d. Plot forecasts from an ARIMA(2,1,2) model with drift and compare these to part a. Remove the constant and see what happens.

e. Plot forecasts from an ARIMA(0,2,1) model with a constant. What happens?

2. Seasonal ARIMA

a. Load `diabetes.rds` from the directory `_data/cleaned`.

b. Perform the following transformation of the column `Cost`: Apply a log transform followed by a seasonal difference. Label the resulting time series `Y`.

c. Apply the KPSS test to `Y`. What is its p-value? What can you conclude about `Y`?

d. Make a time plot of `log(Cost)`. Why does the trend disappear when we consider `Y`?

e. Fit an ARIMA model to `log(Cost)` and report the order of the fitted model.

f. How many fitted parameters are there in the model?

3. Model selection

a. What are the null hypothesis and assumptions of the ADF test?

b. Is it possible for both the ADF and KPSS test applied to a dataset to have large p-values? Explain why or why not.

c. What are the AIC and AICc penalties for the model fitted in Q2?

d. Fit an exponential smoothing model of your choice to `diabetes.rds`. Use `glance()` to view the log likelihood and AICc values of both this model and the ARIMA model from Q2.

e. Can we say which method is a better fit to the data by comparing their log likelihood or AICc? Explain why or why not.

4. Dynamic regression (adapted from Q10.4 in Hyndman & Athanasopoulos)

This exercise concerns `aus_accommodation`: the total quarterly takings from accommodation and the room occupancy level for hotels, motels, and guest houses in Australia, between January 1998 and June 2016. Total quarterly takings are in millions of Australian dollars.

a. Perform inflation adjustment for Takings (using the `CPI` column), creating a new column in the tsibble called `AdjTakings`.

b. For each state, fit a dynamic regression model of `AdjTakings` with seasonal dummy variables, a piecewise linear time trend with one knot at 2008 Q1, and ARIMA errors.

c. What model was fitted for the state of Victoria? Does the time series exhibit constant seasonality?

d. Check that the residuals of the model in c) look like white noise.

e. Use an ARIMA model to forecast CPI, then use these to forecast the (non-adjusted) takings for Victoria until the end of 2017. *Hint: You will have to use the* `new_data` *argument to* `forecast()`.

f. What sources of uncertainty have not been taken into account in the prediction intervals?

5. State space models

Consider the model

$$T_t = T_{t-1} + \epsilon_t$$

$$S_t = 0.8S_{t-4} + \eta_t$$

$$Y_t = T_t + S_t + V_t$$

where $(\epsilon_t) \sim WN(0, \sigma_1^2)$, $\eta_t \sim WN(0, \sigma_2^2)$, and $(V_t) \sim WN(0, \sigma_3^2)$. We observe data $Y_1, \dots, Y_n$.

a. Write the model in the standard (matrix) form of a linear Gaussian state space model.

b. Does $\lim_{t\to\infty} \text{Var}(S_t - \hat{S}_{t|n})$ exist? If so, what is its value?

c. Does $\lim_{t\to\infty} \text{Var}(T_t - \hat{T}_{t|n})$ exist? If so, what is its value?